

was always less than the standard error when only  $x_2$  was taken into account, unless

$$\rho_{13} = 0.$$

We may now prove the similar theorem that when we use three variables,  $x_2, x_3, x_4$ , on which to base the estimate, the standard error will be again decreased, unless

$$\rho_{14} = 0.$$

The condition that  $S(u^2)$ , in our present case, shall be less than  $S(r^2)$  in the last, is, in fact,

$$\left\{ \begin{array}{l} r_{12}^2 + r_{13}^2 + r_{14}^2 - r_{12}^2 r_{34}^2 - r_{23}^2 r_{14}^2 - r_{13}^2 r_{24}^2 \\ - 2(r_{13}r_{14}r_{34} + r_{12}r_{13}r_{23} + r_{12}r_{14}r_{24}) \\ + 2(r_{12}r_{14}r_{23}r_{34} + r_{14}r_{13}r_{24}r_{23} + r_{12}r_{13}r_{24}r_{34}) \end{array} \right\} (1 - r_{23}^2) \\ > (r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23})(1 - r_{23}^2 - r_{24}^2 - r_{34}^2 + 2r_{23}r_{24}r_{34}).$$

This may be finally reduced to—

$$(r_{14} - r_{13}r_{34} - r_{12}r_{24} - r_{14}r_{23}^2 + r_{13}r_{23}r_{24} + r_{12}r_{23}r_{34})^2 > 0,$$

that is

$$\rho_{14}^2 > 0.$$

The treatment of the general case of  $n$  variables, so far as regards obtaining the regressions, is obvious, and it is unnecessary to give it at length.

We can now see that the use of normal regression formulæ is quite legitimate in all cases, so long as the necessary limitations of interpretation are recognised. Bravais'  $r$  always remains a coefficient of correlation. These results I must plead as justification for my use of normal formulæ in two cases\* where the correlation was markedly non-normal.

“Mathematical Contributions to the Theory of Evolution.—On a Form of Spurious Correlation which may arise when Indices are used in the Measurement of Organs.” By KARL PEARSON, F.R.S., University College, London. Received December 29, 1896,—Read February 18, 1897.

(1) If the ratio of two absolute measurements on the same or different organs be taken it is convenient to term this ratio an *index*.

If  $u = f_1(x, y)$  and  $v = f_2(z, y)$  be two functions of the three variables  $x, y, z$ , and these variables be selected at random so that there exists no correlation between  $x, y, y, z$ , or  $z, x$ , there will still be found to

\* ‘Economic Journal,’ Dec., 1895, and Dec., 1896, “On the Correlation of Total Pauperism with Proportion of Out-relief.”

exist correlation between  $u$  and  $v$ . Thus a real danger arises when a statistical biologist attributes the correlation between two functions, like  $u$  and  $v$  to *organic* relationship. The particular case that is likely to occur is when  $u$  and  $v$  are indices with the same denominator for the correlation of indices seems at first sight a very plausible measure of organic correlation.

The difficulty and danger which arise from the use of indices was brought home to me recently in an endeavour to deal with a considerable series of personal equation data. In this case it was convenient to divide the errors made by three observers in estimating a variable quantity by the actual value of the quantity. As a result there appeared a high degree of correlation between three series of absolutely independent judgments. It was some time before I realised that this correlation had nothing to do with the manner of judging, but was a special case of the above principle due to the use of indices.

A further illustration is of the following kind. Select three numbers within certain ranges at random, say  $x, y, z$ , these will be pair and pair uncorrelated. Form the proper fractions  $x/y$  and  $z/y$  for each triplet, and correlation will be found between these indices.

The application of this idea to biology seems of considerable importance. For example, a quantity of bones are taken from an *ossuarium*, and are put together in groups, which are asserted to be those of individual skeletons. To test this a biologist takes the triplet femur, tibia, humerus, and seeks the correlation between the indices *femur/humerus* and *tibia/humerus*. He might reasonably conclude that this correlation marked organic relationship, and believe that the bones had really been put together substantially in their individual grouping. As a matter of fact, since the coefficients of variation for femur, tibia, and humerus are approximately equal, there would be, as we shall see later, a correlation of about 0.4 to 0.5 between these indices had the bones been sorted absolutely at random. I term this a spurious organic correlation, or simply a spurious correlation. I understand by this phrase the amount of correlation which would still exist between the indices, were the absolute lengths on which they depend distributed at random.

It has hitherto been usual to measure the organic correlation of the organs of shrimps, prawns, crabs, &c., by the correlation of indices in which the denominator represents the total body length or total carapace length. Now suppose a table formed of the absolute lengths and the indices of, say, some thousand individuals. Let an "imp" (allied to the Maxwellian demon) redistribute the indices at random, they would then exhibit no correlation; if the corresponding absolute lengths followed along with the indices in the redistribution, they also would exhibit no correlation. Now let us suppose the indices not to have been calculated, but the imp to redistribute the abso-

lute lengths; these would now exhibit no organic correlation, but the indices calculated from this random distribution would have a correlation nearly as high, if not in some cases higher than before. The biologist would be not unlikely to argue that the index correlation of the imp-assorted, but probably, from the vital standpoint, impossible beings was "organic."

As a last illustration, suppose 1000 skeletons obtained by distributing component bones at random. Between none of their bones will these individuals exhibit correlation. Wire the spurious skeletons together and photograph them all, so that their stature in the photographs is the same; the series of photographs, if measured, will show correlation between their parts. It seems to me that the biologist who reduces the parts of an animal to fractions of some one length measured upon it is dealing with a series very much like these photographs. A part of the correlation he discovers between organs is undoubtedly organic, but another part is solely due to the nature of his arithmetic, and as a measure of organic relationship is spurious.

Returning to our problem of the randomly distributed bones, let us suppose the indices femur/humerus and tibia/humerus to have a correlation of 0.45. Now suppose successively 1, 2, 3, 4, &c., per cent. of the bones are assorted in their true groupings, then begins the true organic correlating of the bones. It starts from 0.45, and will alter gradually until 100 per cent. of the bones are truly grouped. The final value may be greater or less than 0.45, but it would seem that 0.45 is a more correct point to measure the organic correlation from than zero. At any rate it appears fairly certain that if a biologist recognised that a perfectly random selection of organs would still lead to a correlation of organ-indices, he would be unlikely to accept index-correlation as a fair measure of the relative intensity of correlation between organs. I shall accordingly define spurious organic correlation as the correlation which will be found between indices, when the absolute values of the organs have been selected purely at random. In estimating relative correlation by the hitherto usual measurement of indices, it seems to me that a statement of the amount of spurious correlation ought always to be made.

(2) *Proposition I.—To find the mean of an index in terms of the means, coefficients of variation, and coefficient of correlation of the two absolute measurements.\**

Let  $x_1, x_2, x_3, x_4$  be the absolute sizes of any four correlated organs;  $m_1, m_2, m_3, m_4$  their mean values;  $\sigma_1, \sigma_2, \sigma_3, \sigma_4$  their standard deviations;

\* In all that follows, unless otherwise stated, the correlation may be of any kind whatever, *i.e.*, the frequencies are not supposed to follow the Gaussian or normal law of error.

$v_1, v_2, v_3, v_4$  their coefficients of variation, i.e.,  $\sigma_1/m_1, \sigma_2/m_2, \sigma_3/m_3, \sigma_4/m_4$  respectively;  $r_{12}, r_{23}, r_{34}, r_{41}, r_{24}, r_{13}$ , the six coefficients of correlation;  $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4$  the deviations of the four organs from their means, i.e.,  $x_1 = m_1 + \epsilon_1, x_2 = m_2 + \epsilon_2, x_3 = m_3 + \epsilon_3, x_4 = m_4 + \epsilon_4$ ;  $i_{13}$  the mean value of the index  $x_1/x_3$ , and  $i_{24}$  the mean value of  $x_2/x_4$ ;  $\Sigma_1, \Sigma_2$  the standard deviations of the indices  $x_1/x_3$  and  $x_2/x_4$  respectively; and  $n$  the total number of groups of organs.

We shall suppose the ratios of the deviations to the mean absolute values of the organs are so small that their cubes may be neglected.

Then

$$\begin{aligned} i_{13} &= \frac{1}{n} S\left(\frac{x_1}{x_3}\right) = \frac{1}{n} \frac{m_1}{m_3} S\left\{\left(1 + \frac{\epsilon_1}{m_1}\right)\left(1 + \frac{\epsilon_3}{m_3}\right)^{-1}\right\} \\ &= \frac{1}{n} \frac{m_1}{m_3} \left\{n + \frac{S(\epsilon_1)}{m_1} - \frac{S(\epsilon_3)}{m_3} - \frac{S(\epsilon_1\epsilon_3)}{m_1 m_3} + \frac{S(\epsilon_3^2)}{m_3^2}\right\}, \end{aligned}$$

if we neglect quantities of the third order in  $\epsilon/m$ . But  $S(\epsilon_1) = S(\epsilon_3) = 0$ ,  $S(\epsilon_1\epsilon_3) = nr_{13}\sigma_1\sigma_3$ , and  $S(\epsilon_3^2) = n\sigma_3^2$ .

Hence:

$$i_{13} = \frac{m_1}{m_3} (1 + v_3^2 - r_{13}v_1v_3) \dots\dots\dots (i).$$

Similarly

$$i_{24} = \frac{m_2}{m_4} (1 + v_4^2 - r_{24}v_2v_4) \dots\dots\dots (ii).$$

Thus we see that the mean of an index is not the ratio of the means of the corresponding absolute measurements, but differs by a quantity depending on the correlation and variation coefficients of the absolute measurements.

(3) *Proposition II.*—To find the standard deviation of an index in terms of the coefficients of variation, and coefficient of correlation of the two absolute measurements.

$$\begin{aligned} n\Sigma_{13}^2 &= S\left(\frac{x_1}{x_3} - i_{13}\right)^2 \\ &= \frac{m_1^2}{m_3^2} S\left\{\left(1 + \frac{\epsilon_1}{m_1}\right)\left(1 + \frac{\epsilon_3}{m_3}\right)^{-1} - (1 + v_3^2 - r_{13}v_1v_3)\right\}^2 \\ &= \frac{m_1^2}{m_3^2} \left\{S\left(\frac{\epsilon_1}{m_1} - \frac{\epsilon_3}{m_3} + \text{square terms}\right)^2\right\} \\ &= i_{13}^2 (nv_1^2 + nv_3^2 - 2nr_{13}v_1v_3), \end{aligned}$$

if we neglect cubic terms.

$$\therefore \Sigma_{13} = i_{13} \sqrt{(v_1^2 + v_3^2 - 2r_{13}v_1v_3)} \dots\dots\dots (iii).$$

(4) *Proposition III.*—To find the coefficient of correlation of two indices in terms of the coefficients of correlation of the four absolute measurements and their coefficients of variation.

Let  $x_1/x_3$  and  $x_2/x_4$  be the two indices.

$$\begin{aligned} n\rho\Sigma_{13}\Sigma_{24} &= S\left(\frac{x_1}{x_3}-i_{13}\right)\left(\frac{x_2}{x_4}-i_{24}\right) \\ &= \frac{m_1m_2}{m_3m_4}S\left(1+\frac{\epsilon_1}{m_1}-\frac{\epsilon_3}{m_3}-\frac{\epsilon_1\epsilon_3}{m_1m_3}+\frac{\epsilon_3^2}{m_3^2}-1-v_3^2+r_{13}v_1v_3\right) \\ &\quad \times \left(1+\frac{\epsilon_2}{m_2}-\frac{\epsilon_4}{m_4}-\frac{\epsilon_2\epsilon_4}{m_2m_4}+\frac{\epsilon_4^2}{m_4^2}-1-v_4^2+r_{14}v_1v_4\right) \\ &= i_{13}i_{24}S\left(\frac{\epsilon_1}{m_1}-\frac{\epsilon_3}{m_3}\right)\left(\frac{\epsilon_2}{m_2}-\frac{\epsilon_4}{m_4}\right), \end{aligned}$$

if we neglect terms of the cubic order.

$$\therefore \rho\Sigma_{13}\Sigma_{24} = i_{13}i_{24}(r_{12}v_1v_2-r_{14}v_1v_4-r_{23}v_2v_3+r_{34}v_3v_4).$$

Hence, finally,

$$\rho = \frac{r_{12}v_1v_2-r_{14}v_1v_4-r_{23}v_2v_3+r_{34}v_3v_4}{\sqrt{v_1^2+v_3^2-2r_{13}v_1v_3}\sqrt{v_2^2+v_4^2-2r_{24}v_2v_4}} \dots\dots\dots (iv).$$

(5) Thus we have expressed  $\rho$  in terms of the four coefficients of correlation and the four coefficients of variation of the absolute measurements which form the indices.

We may draw the following conclusions :

(i.) The correlation between two indices will always vanish when the four absolute measurements forming the indices are quite uncorrelated.

(ii.) If two of the organs are perfectly correlated, let us say made identical: for example, the third and fourth, so that  $r_{34} = 1$ , and  $v_3 = v_4$ , we find

$$\rho = \frac{r_{12}v_1v_2-r_{13}v_1v_3-r_{23}v_2v_3+v_3^2}{\sqrt{v_1^2+v_3^2-2r_{13}v_1v_3}\sqrt{v_2^2+v_3^2-2r_{23}v_2v_3}} \dots\dots\dots (v).$$

This is the coefficient of correlation between two indices with the same denominator ( $x_1/x_3$  and  $x_2/x_3$ ).

The value of  $\rho$  in (v) does not vanish if the remaining organs be quite uncorrelated, i.e.,  $r_{12} = r_{13} = r_{23} = 0$ . In this case

$$\rho_0 = \frac{v_3^2}{\sqrt{v_1^2+v_3^2}\sqrt{v_2^2+v_3^2}} \dots\dots\dots (vi).$$

This is the measure of the spurious correlation. For the special

case in which the coefficients of variation are all the same,  $\rho_0 = 0.5$ . When the absolute sizes of organs are very feebly correlated, then in most cases there will be a considerable correlation of indices.

*Example (a).* Suppose three organs,  $x_1$ ,  $x_2$ , and  $x_3$  to have sensibly equal coefficients of variation, and that the correlation of  $x_1$  and  $x_2 = r_{12} = r$  and of  $x_1$  and  $x_3$ , as well as of  $x_2$  and  $x_3 = r'$ .

Then :

$$\begin{aligned}\rho &= 0.5 \times \frac{1+r-2r'}{1-r'} \\ &= 0.5 + 0.5 \frac{r-r'}{1-r'}.\end{aligned}$$

This formula illustrates well in a specially simple case how the correlation in the indices diverges from the spurious value 0.5, as we alter  $r$  and  $r'$  from zero, *i.e.*, as we introduce organic correlation. According as  $r$ , the correlation of the numerators, is greater or less than  $r'$ , the correlation of the numerator with the denominator, the actual index correlation can be greater or less than the spurious value.

*Example (b).* If  $z_1$ ,  $z_2$  be the indices, then in the case of normal correlation the contour lines of the correlation surface for the indices are given by

$$\frac{z_1^2}{\Sigma_1^2(1-\rho^2)} - \frac{2\rho z_1 z_2}{\Sigma_1 \Sigma_2(1-\rho^2)} + \frac{z_2^2}{\Sigma_2^2(1-\rho^2)} = \text{constant},$$

where  $\Sigma_1$ ,  $\Sigma_2$ , and  $\rho$  are given by (iii) and (iv) above.

The contour lines of a surface of spurious index correlation are given by

$$\frac{z_1^2}{\sigma_1^2} \left( 1 + \frac{v_3^2}{v_2^2} \right) - 2 \frac{z_1 z_2}{\sigma_1 \sigma_2} \frac{v_3^2}{v_1 v_2} + \frac{z_2^2}{\sigma_2^2} \left( 1 + \frac{v_3^2}{v_1^2} \right) = \text{constant},$$

while the uncorrelated distribution of the numerators  $x_1$  and  $x_2$  is given by the contours,

$$x_1^2/\sigma_1^2 + x_2^2/\sigma_2^2 = \text{constant}.$$

We are thus able to mark the growth of the spurious correlation as we increase  $v_3$  from zero ; we see the axes of the ellipses diminishing and their directions beginning to rotate.

*Example (c).* To find the spurious correlation between the two chief cephalic indices.

I have calculated the following results from the measurements made on 100 "Alt-bayerisch" ♂ skulls, by Professor J. Ranke. See his 'Anthropologie der Bayern,' Bd. i, Kapitel v, S. 194.

Breadth of skull :*	$m_1 = 150\cdot47,$	$\sigma_1 = 5\cdot8488,$	$v_1 = 3\cdot8871.$
Height of skull :	$m_2 = 133\cdot78,$	$\sigma_2 = 4\cdot6761,$	$v_2 = 3\cdot4954.$
Length of skull :	$m_3 = 180\cdot58,$	$\sigma_3 = 5\cdot8441,$	$v_3 = 3\cdot2363.$
Cephalic index, B/L :	$i_{13} = 83\cdot41,$	$\Sigma_{13} = 3\cdot5794,$	$V_{13} = 4\cdot2913.$
Cephalic index, H/L :	$i_{23} = 74\cdot23,$	$\Sigma_{23} = 3\cdot6305,$	$V_{23} = 4\cdot8909.$
Cephalic index, H/B :	$i_{21} = 89\cdot12,$	$\Sigma_{21} = 4\cdot1752,$	$V_{21} = 4\cdot6849.$

The coefficients of correlation may at once be deduced :

$$\text{Breadth and length : } r_{13} = (v_1^2 + v_3^2 - V_{13}^2)/(2v_2v_3) = 0\cdot2849.$$

$$\text{Height and length : } r_{23} = (v_2^2 + v_3^2 - V_{23}^2)/(2v_2v_3) = -0\cdot0543.$$

$$\text{Height and breadth : } r_{21} = (v_2^2 + v_1^2 - V_{21}^2)/(2v_1v_2) = 0\cdot1243.$$

This is the first table, so far as I am aware, that has been published of the variation and correlation of the three chief cephalic lengths.† It shows us that there is not at all a close correlation between these chief dimensions of the skull, and that a small compensating factor for size is to be sought in the correlation of height and length, *i.e.*, while a broad skull is probably a long skull and also a high skull, a high skull will probably be a short skull, and a low skull a long skull.

Without substituting the values of  $v_1, v_2, v_3, r_{12}, r_{13}, r_{23}$  in (v), we can find  $\rho$ , or the correlation between breadth/length and height/length indices from :

$$\rho = (V_{13}^2 + V_{23}^2 - V_{12}^2)/(2V_{13}V_{23}).$$

This follows at once from the general theorem given in my memoir on "Regression, Panmixia, and Heredity," 'Phil. Trans.,' vol. 187, A, p. 279, or by substitution of the above values of  $r_{12}, r_{13}, r_{23}$  in (v), we find :

$$\rho = 0\cdot4857.$$

If we calculate from (vi) the correlation between the same cephalic indices on the hypothesis that their heights, breadths and lengths are distributed at random, *i.e.*, that our "imp" has constructed a number of arbitrary and spurious skulls from Professor Ranke's measurements, we find :

$$\rho_0 = 0\cdot4008.$$

It seems to me that a quite erroneous impression would be formed of the organic correlation of the human skull, did we judge it by the magnitude of the correlation coefficient (0·4857) for the two chief

\* All the absolute measures given are in millimetres, and the coefficients of variation are *percentage* variations, *i.e.*, they must be divided by 100 before being used in formulæ (i), (ii), and (iii).

† I hope later to treat correlation in man with reference to race, sex, and organ, as I have treated variation.

cephalic indices, for no less than 0·4008 of this would remain, if we destroyed all organic relationship between the lengths on which these indices are based.

*Example (d). To find the spurious correlation between the indices femur/humerus and femur/tibia.*

The following results have been calculated\* from measurements made by Koganei on Aino skeletons. (See 'Mittheilungen aus der medicinischen Facultät der K. J. Universität, Tokio,' Bd. I. Tables.)

I have kept the sexes apart although there are but few of each.

♂ Skeletons. Number = 40 to 44. Measurements in centimetres.

Femur, F:  $m_1 = 40\cdot845$ ,  $\sigma_1 = 1\cdot957$ ,  $v_1 = 4\cdot792$ .

Tibia, T:  $m_2 = 31\cdot740$ ,  $\sigma_2 = 1\cdot577$ ,  $v_2 = 4\cdot970$ .

Humerus, H:  $m_3 = 29\cdot593$ ,  $\sigma_3 = 1\cdot337$ ,  $v_3 = 4\cdot517$ .

The following coefficients of correlation were calculated directly:

Femur and tibia:  $r_{12} = 0\cdot8266$ .

Femur and humerus:  $r_{13} = 0\cdot8585$ .

Tibia and humerus:  $r_{23} = 0\cdot7447$ .

From these were deduced by the formulæ of this paper†:—

Index, F/T:  $i_{12} = 128\cdot75$ ,  $\Sigma_{12} = 3\cdot7075$ ,  $V_{12} = 2\cdot8795$ .

Index, F/H:  $i_{13} = 137\cdot92$ ,  $\Sigma_{13} = 3\cdot4084$ ,  $V_{13} = 2\cdot4714$ .

Index, T/H:  $i_{23} = 107\cdot02$ ,  $\Sigma_{23} = 3\cdot6675$ ,  $V_{23} = 3\cdot4271$ .

Hence we find for the correlation of the indices F/H and T/H:

$$\rho = 0\cdot5644.$$

But the spurious correlation, if the bones had been grouped at random would have been

$$\rho_0 = 0\cdot4557.$$

\* I have to thank Miss Alice Lee for a considerable part of the arithmetic work of this example.

† The values for the indices are not in absolute agreement with those to be deduced from the lengths, for it was not always possible to use the same skeleton for femur and humerus as for tibia and humerus, *i.e.*, sometimes one or other bone was missing. For the same reason, the constants for the absolute lengths do not agree entirely with those given for Ainos in my paper on "Variation in Man and Woman" in 'The Chances of Death and other Studies in Evolution,' vol. I, p. 303), for the simple reason that I there used every available bone, and not every available *pair*, as here.



Tabulating the corresponding quantities for the other sex we find

♀ Skeletons. Number = 22 to 24. Measurements in centimetres.

Femur, F :	$m_1 = 38.075,$	$\sigma_1 = 1.494,$	$v_1 = 3.924.$
Tibia, T :	$m_2 = 29.800,$	$\sigma_2 = 1.576,$	$v_2 = 5.289.$
Humerus, H :	$m_3 = 27.565,$	$\sigma_3 = 1.109,$	$v_3 = 4.022.$
Femur and tibia :	$r_{12} = 0.8457.$		
Femur and humerus :	$r_{13} = 0.8922.$		
Tibia and humerus :	$r_{23} = 0.7277.$		
Index, F/T :	$i_{12} = 127.90,$	$\Sigma_{12} = 3.8937,$	$V_{12} = 3.0444.$
Index, F/H :	$i_{13} = 138.37,$	$\Sigma_{13} = 2.6930,$	$V_{13} = 1.9462.$
Index, T/H :	$i_{23} = 108.36,$	$\Sigma_{23} = 4.1022,$	$V_{23} = 3.7857.$
	$\rho = 0.6006.$		
	$\rho_0 = 0.3904.$		

Hence we may conclude as follows :

(i) The absolute lengths of the long bones differ from those of the skull in being very closely correlated.

(ii) The use of indices for the long bones would appear to minimise, rather than, as in the case of the skull, to exaggerate this correlation.

(iii) If we measure, however, organic correlation of the indices by  $\rho - \rho_0$ , we shall find index correlation less than absolute length correlation for both long bones and skull, and in both cases the former comparatively small as compared with the latter.

(iv) The results for the 24 female skeletons, although based on but few data, serve on the whole to confirm the male results.\*

(6.) From the above examples it will be seen that the method, which judges of the intensity of organic correlation by the reduction of all absolute measures to indices, the denominators of which are some *one* absolute measurement, is not free from obscurity ; for this method would give the major portion of the observed index correlation had the parts of the animal been thrown together entirely at random, *i.e.*, if there were no organic correlation at all. The following additional remarks may be of interest. The results (iv)—(vi) show us that the correlation coefficients of indices are functions, not only of the correlation coefficients of absolute measurements, but also of the coefficients of variation of the latter measurements. Hence,

\* The fact that the male is more variable in height-sitting, in femur, and in tibia than the female, while she appears to be more variable than he is in stature, led me to prophesy, in my paper on "Variation in Man and Woman," that the female would be found to be more closely correlated in the bones forming stature than the male. This appears to be the case for the femur and tibia of Ainos.

unless the coefficients of variation be constant for local races, it is impossible that the coefficients of correlation can be constant for indices. In other words, the hypothesis of the constancy for local races of correlation, and that of the constancy for local races of variation, stand on exactly the same footing.

The conclusions of this paper although applied to organic correlation are equally valid so far as concerns the use of indices in judging the correlation of either physical or economic phenomena. It was, indeed, a difficulty arising from my discussion of personal judgments—a spurious correlation between the judgments of different observers—which first drew my attention to the matter.

*Note, January 13, 1897.*—The result described by Professor Pearson evidently affects the value of the correlation coefficients determined by me in *Crangon* and *Carcinus* ('Roy. Soc. Proc.,' vols. 51 and 54), because I have always expressed the size of the organs measured in terms of body length.

In order to show the effect of this, I have lately performed, at Professor Pearson's suggestion, the following experiment: It happens that my measures of Plymouth shrimps are recorded in a book, in the order in which they were measured, and therefore at random as regards carapace length or other characters. I constructed from these records 420 "spurious" shrimps, in the following way: the total length of the first shrimp in the book was associated with the carapace length of the tenth shrimp and the "post-spinous length" of the twentieth, and so throughout. Evidently these three measures were associated at random, and we might expect that these spurious shrimps would show no organic correlation; but when the carapace lengths and "post-spinous lengths" of these spurious shrimps were divided by the body length, and the correlation between the resulting indices was determined, the value of  $r$  was found to be 0.38, the value for real shrimps being 0.81, or the correlation due to the use of indices forms 47 per cent. of the observed value.

W. F. R. WELDON.

"Note to the Memoir by Professor Karl Pearson, F.R.S., on Spurious Correlation." By FRANCIS GALTON, F.R.S. Received January 4,—Read February 18, 1897.

I send this note to serve as a kind of appendix to the memoir of Professor K. Pearson, believing that it may be useful in enabling others to realise the genesis of spurious correlation. It is important though rather difficult to do so, because the results arrived at in the memoir, which are of serious interest to practical statisticians, have at first sight a somewhat paradoxical appearance.